# Spectroscopy of DNS Update Traffic

Andre Broido, Evi Nemeth, kc claffy
CAIDA, San Diego Supercomputer Center,
University of California, San Diego
E-mail: {broido, evi, kc}@caida.org*

## ABSTRACT

We study attempts to dynamically update DNS records for private (RFC1918) addresses, by analyzing the frequency spectrum of updates observed at an authoritative nameserver for these addresses. Using a discrete autocorrelation algorithm we found that updates series have periods of 60 or 75 minutes, which we identified as default settings of out-of-the-box Microsoft Windows 2000 and XP DNS software.

## General Terms

Measurement

## Categories and Subject Descriptors

C2.2 [**Computer-Communication Networks:**]: Network Protocols

## Keywords

DNS, root servers, RFC1918 addresses, spectroscopy

## 1. INTRODUCTION

In this paper we analyze spurious machine-generated traffic in the worldwide Domain Name System (DNS) – attempts to update address-to-hostname mappings in nameservers at the top of the DNS hierarchy.

Most of the DNS traffic we observed dealt with so called private, or RFC1918 addresses. We discovered that a large portion of these updates are caused by the default configuration of the DHCP/DNS servers shipped with Microsoft systems. This server software sends periodic updates with frequencies that we found with spectral analysis and confirmed by laboratory experiment and vendor documentation. This (mis)configuration is so widespread that patterns of Internet access by end users are reflected in the pulsations of the flow of DNS updates. The resulting traffic is not only a waste of global Internet resources, but also raises security, privacy and intellectual property questions of its own.

Since mid-2002 almost all RFC1918 update traffic is deflected from the roots to *blackhole servers*. These servers do not solve the problem but at least protect the roots from misguided traffic. The creation of worldwide system of authoritative servers for RFC1918 addresses[1] as proposed by Paul Vixie proved feasibility of anycast routing for building global infrastructures, and enabled diversification of root servers' deployment. An unsolved problem is how to protect servers from load swings when route preferences suddenly change: a DNS question becomes a routing challenge.

The analysis we present here extends work on measurement, performance and placement of DNS root servers [1] In particular, [1] [2] discuss the ubiquity of DNS misconfiguration in queries at the root servers. For a complementary view on the root server load problem, we recommend [3]. The full version of this paper is also available [4].

The identification techniques that we call *network spectroscopy* are based on patterns of delay quantization. They apply to a variety of other setups [5]. We have used them for bitrate estimation and broadband source recognition [6], and we are currently testing them on BGP updates.

## 2. RESULTS

We present here analyses of two datasets, D1 (May 28-June 04, 2002, 98 M updates, 1.2 M source IP addresses) and D2 (July 04-30, 304 M updates, 2.4 M IPs.)

A preliminary analysis of possible causes of the RFC1918 updates phenomenon revealed that:

1. The volume of updates surges sharply at local midnight for each time zone with a large population of Internet users. Daily/weekly update rate variations are consistent with common patterns of human activity.

2. The majority (60%) of updates are from sources that send them constantly (Fig.1a.) Most of updates are from medium-rate contributors rather than from mice or elephants (Fig.1b.) Average update rates have two spikes near 1 and 2.4 updates/hour (Fig.2a) for both source IP and update weights, and a hump at 1 update per 1-3 days (IP weight.)

3. Most source IP addresses are those of home and small business users connected to the Internet via cable, DSL or phone-based Internet providers. Academic, corporate and backbone networks contribute a minor amount of updates.

Our observations indicate that DNS updates come from computers owned by individuals, not organizations. The majority of them use the software with default vendor settings. It is natural to assume (cf. observation (2)), that persistent update generation is the default behavior of Mi-

*Support for this work is provided by the Defense Advanced Research Project Agency (DARPA), through its NGI (N66001-98-2-8922) and NMS (N66001-01-1-8909) programs, and by the National Science Foundation (NSF NCR-9711092).


*SIGMETRICS'03,* June 10–14, 2003, San Diego, California, USA.
ACM 1-58113-664-1/03/0006.

[1]Blackhole servers use addresses in 192.175.48.0/24. This prefix is announced by 14 or more ASes, mostly with origin AS 112 (see `www.as112.net` for details.)

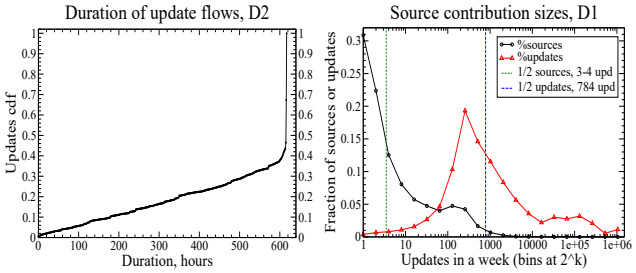


**Figure 1: a) Update flow duration cdf b) Mice, mules and elephants**

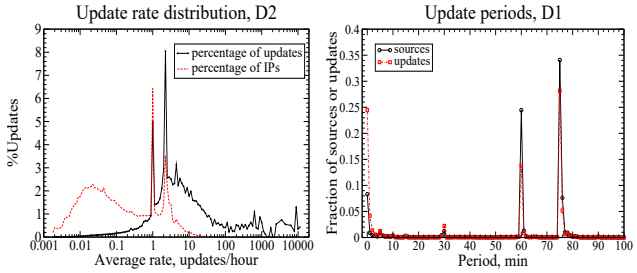


**Figure 2: a) Update rates b) Update periods**

crosoft's DNS implementation. To find out what causes the periodic update traffic, we took the following steps:

1. Analyzed interarrival times, identifying two narrow spikes using a discrete autocorrelation function (see below.) The prevalent periods were found to reflect one update per hour and 3 updates per 75 minutes.

2. Set up laboratory experiment with off-the-shelf software confirming that Windows (2000 and XP) DHCP/DNS servers send periodic DNS updates.

3. Found Microsoft documentation describing their DNS update implementation with observed periods as the default behavior for their operating systems.

## 2.1 Update periods

The update data contains interleaved sequences sent on behalf of several local hosts that can join and leave the private network. Together with the occasional missing or extra updates, these variations require a robust algorithm to distill. That is why we matched binary orders of magnitude rather than numeric values. Our discrete autocorrelation algorithm finds the fraction of periodic updates as follows:

1. For any source with $n \geq 10$ updates, take binary logarithms $b_i = \lfloor \log_2 d_i \rfloor$ of interarrival delays $d_i, i = 1, \ldots, n$.

2. For each shift $k$ of the sequence $b_i$ by $k = 1, \ldots, 30$, count the number of positions $r$ in which $b_i = b_{i+k}$.

3. Find the lag $l > 0$ which maximizes match count $r$; discard the source if $r < 0.1n$.

4. Find the longest interval $g \leq i \leq g'$ with $b_i = b_{i+l}$.

5. Extract $l$ interarrival times $d_g, \ldots, d_{g+l-1}$. Take the sum $\sum_{i=g}^{g+l-1} d_i$ as the period's estimate.

This algorithm was able to disambiguate interleaved sequences (Fig.2b). See [4] for NATted hosts counts (cf. [7].)

## 2.2 Microsoft documentation

The Microsoft documentation admits to periodic updates and spikes at local midnight. The latter come from the NetLogon service trying to register the forward and reverse DNS mappings every 24 hours [8].[2]

The 75 minute period is a sum of 5, 10, and 60 minute timeouts (Windows 2000 DNS Whitepaper [9].)

Before the introduction of blackhole servers, the leakage of private addresses to root servers was caused by *devolution* (removal of higher-level labels from domain) when trying to locate server authoritative for an address-to-name map (PTR record.) Since e.g. `168.192.in-addr.arpa` had no authoritative server, devolution proceeded to the second-level domain. The vendor's algorithm missed the fact that authority for `in-addr.arpa` is vested in the root servers.

[2] "By default, DNS records are re-registered dynamically and periodically every 24 hours by Windows 2000 Professional and every 1 hour by Windows 2000 Server and Windows 2000 Advanced Server" [8]. "By default, the DHCP server updates the PTR resource record" [9]. This applies to all IP addresses served by DHCP, private and globally unique.

# 3. CONCLUSIONS

We have demonstrated that the majority of periodic updates derives from Windows 2000 and Windows XP. Prior to the deployment of the AS112 authoritative servers for RFC1918 address space (Spring 2002), Microsoft-based machines with private addresses tried to update the DNS root servers, which can be compared to a massive DDoS attack.

We conclude that Microsoft must change the default configuration so that dynamic DNS updates are disabled and user configuration, or lack thereof, does not enable RFC1918-related traffic to propagate beyond the local subnet.

More generally we consider this study a compelling example of why software and setups affecting stability of the Internet's infrastructure must be designed with careful attention to potential effects of engineering decisions. Indeed the current state of desktop software poses a burden on, if not threat to, the robustness of the global Internet.

## 3.1 Acknowledgements

Many thanks to Paul Vixie, Peter Losher, Brian Kantor, Piet Barber, Cricket Liu, Ryan King, Tom Guptill, Nevil Brownlee, Marina Fomenkov, Ken Keys and IPAM UCLA. The feedback from the IETF and NANOG participants, and from SIGMETRICS reviewers was also highly appreciated.

# 4. REFERENCES

[1] N.Brownlee, kc claffy, and E.Nemeth, "DNS Measurements at a Root Server," Globecom 2001.

[2] D.Wessels and M.Fomenkov, "Wow, that's a lot of packets," in *PAM*, Apr 2003.

[3] R. Liston, S.Srinivasan, and E.Zegura, "Diversity in DNS Performance Measures," in *IMW*, Nov 2002.

[4] A. Broido, E. Nemeth, and kc claffy, "Spectroscopy of private DNS update sources," `www.caida.org`.

[5] D.Katabi and Ch.Blake, "Inferring congestion sharing and link characteristics from packet interarrival times," MIT LCS Technical Report, 2001.

[6] A. Broido, R. King, E. Nemeth, and kc claffy, "Radon spectroscopy of inter-packet delay," in *HSN*, Mar 2003.

[7] Steven M. Bellovin, "A Technique for Counting NATted Hosts," in *IMW*, Nov 2002.

[8] "How to Enable/Disable Windows 2000 Dynamic DNS Registrations," Microsoft Knowledge Base Q246804.

[9] "Windows 2000 DNS Whitepaper," http://www.microsoft.com/windows2000/docs/w2kdns.doc.